Behavioral/Cognitive

The Rapid Emergence of Musical Pitch Structure in Human Cortex

Narayan Sankaran,[1,2] **Thomas A. Carlson,**[2,4] **and William Forde Thompson**[2,3]
[1]Auditory Neuroscience Laboratory, School of Medical Sciences, University of Sydney, Sydney, 2006 New South Wales, Australia, [2]Australian Research Council Centre of Excellence in Cognition and Its Disorders, Macquarie University, Sydney, 2109 New South Wales, Australia, [3]Department of Psychology, Macquarie University, Sydney, 2109 New South Wales, Australia, and [4]School of Psychology, University of Sydney, Sydney, 2006 New South Wales, Australia

In tonal music, continuous acoustic waveforms are mapped onto discrete, hierarchically arranged, internal representations of pitch. To examine the neural dynamics underlying this transformation, we presented male and female human listeners with tones embedded within a Western tonal context while recording their cortical activity using magnetoencephalography. Machine learning classifiers were then trained to decode different tones from their underlying neural activation patterns at each peristimulus time sample, providing a dynamic measure of their dissimilarity in cortex. Comparing the time-varying dissimilarity between tones with the predictions of acoustic and perceptual models, we observed a temporal evolution in the brain's representational structure. Whereas initial dissimilarities mirrored their fundamental-frequency separation, dissimilarities beyond 200 ms reflected the perceptual status of each tone within the tonal hierarchy of Western music. These effects occurred regardless of stimulus regularities within the context or whether listeners were engaged in a task requiring explicit pitch analysis. Lastly, patterns of cortical activity that discriminated between tones became increasingly stable in time as the information coded by those patterns transitioned from low-to-high level properties. Current results reveal the dynamics with which the complex perceptual structure of Western tonal music emerges in cortex at the timescale of an individual tone.

Key words: magnetoencephalography; multivariate pattern analysis; music perception; neural decoding; pitch perception; representational dynamics

Significance Statement

Little is understood about how the brain transforms an acoustic waveform into the complex perceptual structure of musical pitch. Applying neural decoding techniques to the cortical activity of human subjects engaged in music listening, we measured the dynamics of information processing in the brain on a moment-to-moment basis as subjects heard each tone. In the first 200 ms after onset, transient patterns of neural activity coded the fundamental frequency of tones. Subsequently, a period emerged during which more temporally stable activation patterns coded the perceptual status of each tone within the "tonal hierarchy" of Western music. Our results provide a crucial link between the complex perceptual structure of tonal music and the underlying neural dynamics from which it emerges.

Introduction

We continuously and effortlessly extract meaning from the sonic world around us. This meaning emerges when sensory information from the external environment interacts with internally stored domain-specific knowledge. Tonal music, with its formal abstract pitch structure, provides an ideal domain for examining such an interaction (Lerdahl, 1992). In tonal systems the world over, pitch is arranged hierarchically. Depending on the prevailing musical key, certain pitch classes occur more frequently, occupy positions of melodic, harmonic, and rhythmic prominence (Vos and Troost, 1989), and have greater perceived stability (Krumhansl and Shepard, 1979; Krumhansl and Kessler, 1982).

Received June 16, 2019; revised Jan. 6, 2020; accepted Jan. 7, 2020.

Author contributions: N.S., T.A.C., and W.F.T. designed research; N.S. performed research; N.S. contributed unpublished reagents/analytic tools; N.S., T.A.C., and W.F.T. analyzed data; N.S. wrote the first draft of the paper; N.S., T.A.C., and W.F.T. edited the paper; N.S. wrote the paper.

This work was supported by Australian Research Council Centre of Excellence in Cognition and Its Disorders Grant CE110001021, and Australian Research Council Grant FT120100816 Future Fellowship to T.A.C. We thank Yanan Sun for assistance in data collection.

The authors declare no competing financial interests.

Correspondence should be addressed to Narayan Sankaran at narayan.sankaran@ucsf.edu.

https://doi.org/10.1523/JNEUROSCI.1399-19.2020

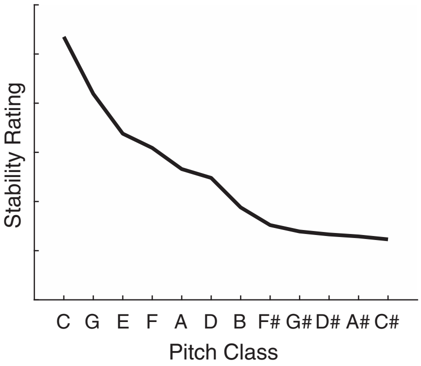

Figure 1. The STH of perceived stability. Ratings of stability assigned to each pitch class within the Western key of C major, as reported by Krumhansl and Kessler (1982). The first scale degree C is maximally stable and heads the hierarchy. Following are the fifth and third scale degrees (G and E, respectively), the other scale tones (D, F, G, A, B), and finally the nonscale or "out-of-key" tones (C#, D#, F#, G#, A#).

This profile of stability has been experimentally quantified for Western tonal music as the standard tonal hierarchy (STH) and forms a cornerstone of music perception (Fig. 1).

Despite its role as the principle schema underlying the organization of tonal music, the neural substrates supporting the STH remain poorly understood. After core auditory areas extract basic frequency information from the acoustic signal, a representation of complex pitch is thought to emerge in secondary auditory regions (Zatorre et al., 1994; Griffiths et al., 1998; Wessinger et al., 2001; Hall et al., 2002; Patterson et al., 2002). How does this isolated sensory representation then acquire the perceived attributes of musical pitch? To do so, the surrounding musical context must be integrated, recruiting cortical populations that reflect prior knowledge of tonal structure. Both lesion and neuroimaging studies have identified regions implicated in the processing of both melodic (Lee et al., 2011) and harmonic (Klein and Zatorre, 2011; Fedorenko et al., 2012; Foo et al., 2016) structure, while electrophysiological research has identified cortical response components sensitive to the hierarchical status of evoking tones (Brattico et al., 2006; Krohn et al., 2007). More recently, Sankaran et al. (2018) showed that tones differing only in their hierarchical status evoked separable patterns of neural activity, suggesting that measurable activity in cortex may code specific information detailing the structure of the tonal hierarchy. Despite these advances, empirical work is yet to fully map the neural representational space of musical pitch and explicitly test the predictions of perceptual and music-theoretic models. The current study therefore evaluated two major questions: first, do cortical populations encode musical pitch in a manner that precipitates the organization of the STH? And second, what are the representational dynamics underlying the emergence of such a perceptual representation from lower-level afferent information?

To address these questions, we recorded cortical activity using MEG as musically trained subjects listened to different pitch classes within a tonal context. We applied multivariate pattern analysis (MVPA) (Haxby et al., 2014) to decode the identity of tones from their corresponding MEG activity, and the accuracy with which classifiers discriminated between the response patterns elicited by two tones provided a proxy for their dissimilarity in the cortex. Moreover, MVPA was applied using a sliding time window, enabling us to track the dynamics of the evolving cortical representation. Comparing these time-varying neural dissimilarities with the predictions of several models of pitch, we quantified the underlying information held in cortical population codes at any given moment along various sensory and perceptual stimulus dimensions.

Materials and Methods

Participants. Eighteen subjects (7 male) with a minimum of 5 years of formal music training (mean = 11.9 years; SD = 3.8 years) were recruited through the Sydney Conservatorium of Music and Macquarie University to partake in the study. All subjects reported having no known hearing loss or brain abnormalities and did not possess absolute pitch. The study was approved beforehand by the Human Research Ethics Committee at Macquarie University (reference #5201300804), and all methods were performed in accordance with the stated guidelines. Informed consent was obtained before testing, after all experimental details and potential risks were explained.

Apparatus. Data were collected with a whole-head MEG system (model PQ1160R-N2, KIT) consisting of 160 coaxial first-order gradiometers with a 50 mm baseline (Kado et al., 1999; Uehara et al., 2003). Before recording, each participant's head shape was measured with a pen digitizer (Polhemus, Fastrack), and the positions of five marker coils on the surface of the scalp were registered. During recording, MEG data were bandpass filtered online from 0.1 to 200 Hz using first-order RC filters and digitized at 1000 Hz. Participants were in a supine position situated within a magnetically shielded room containing the MEG sensors. During experimental trials, participants were instructed to gaze at a fixation cross. Both the fixation cross and experimental instructions were projected by an LCD back projection system (InFocus) to a screen located above the participant at a viewing distance of 113 cm. Sound stimuli were delivered via Etymonic ER-30 insert headphones at 44.1 kHz.

Stimuli. Trials consisted of a tonal context followed by 1 of 12 different tones (hereafter referred to as probe tones). All stimuli were piano tones sampled at 44.1 kHz using a virtual instrument plugin in Max/MSP (Cycling '74). Probe tones and chords were 650 ms in duration (95 bpm). Before testing, all probe tones were passed through a time-varying loudness model (Glasberg and Moore, 2002) and normalized for differences in perceived loudness. To achieve this, the maximum short-term loudness (STL_{max}) of each tone was computed. We used STL_{max} as a measure of perceived loudness as it operates at timescales comparable to individual note processing (Thwaites et al., 2016). We then normalized the loudness of all tones to their average STL_{max}. Modifications in level resulting from this procedure did not exceed 3 phons. Probe tones spanned the chromatic range between F#3 (185 Hz) and F4 (349 Hz), such that their average semitone distance to the preceding context was minimized. The tonal context consisted of four major chords written in four-part harmony outlining an I-IV-V-I harmonic progression in the key of C major. To prevent sensory processing of the context from contaminating evoked responses to probe tones, a silent period of one beat (i.e., 650 ms) separated the two.

Experimental design. MEG activity was recorded from subjects as they listened to the above trials. Probe tones were presented in random order without repeats across trials. To ensure participants were attending to stimuli (Loui et al., 2005), participants judged whether the probe tone on each trial was "in-key" or "out-of-key," registering their response only after stimulus offset by pressing one of two buttons. Participants used their left and right thumbs to register the two respective responses, and the mapping of in-key/out-of-key to left/right button was switched every two blocks to control for the potential effects of motor activity. No trial-by-trial feedback was provided during the MEG recording. On average, subjects responded correctly on 78% of the trials (SD = 16.3%). All trials, including those with incorrect responses, were included in the analysis of MEG recordings (Vanrullen, 2011). Intertrial intervals were randomized between 0.5 and 1 s. Before testing, subjects completed a training session consisting of 20 trials with an identical behavioral task to that of the main experiment. Feedback was provided after each training trial, and the experimenter ensured that subjects could perform the task (using a threshold of ≥75% correct) before proceeding to the MEG recording session. Each participant's MEG data were collected in a single hour-long session. The total experiment comprised 672 trials, yielding 56 observations of each of the 12 probe tones. Testing was divided into 8 blocks, each comprised of 84 trials and separated by 1 min breaks.

Analysis

MEG preprocessing. All preprocessing and subsequent analyses of MEG data were performed in MATLAB (The MathWorks). Neural epochs were extracted from 100 ms before to 1000 ms after onset of probe tones. Data were downsampled to 100 Hz with a low-pass Chebyshev Type 1 filter. Next, principal components analysis (PCA) was applied to the dataset of each participant using all MEG sensor channels as features. Principle components that cumulatively explained 99% of the variance were retained. On average, PCA reduced the dimensionality of datasets from 160 channels to 28 principle components (SD = 5.4). PCA has been found to be an efficient preprocessing step for optimizing data for MEG decoding analyses (Grootswagers et al., 2017). In a single step, PCA reduces the dimensionality of the data and obviates the need for additional artifact rejection or denoising procedures, as classifiers can learn to suppress nuisance variables isolated by PCA (e.g., eye-blinks and environmental noise).

MVPA. To measure the neural dissimilarity between two given probe tones, binary classifiers attempted to decode the identity of tones from the participant's recorded brain activity (Haxby et al., 2014). Before classification, we averaged responses of two exemplars within each class to boost the signal-to-noise ratio of classification. We used a naive Bayes implementation of linear discriminate analysis (Duda et al., 2012) to perform classification for each pairwise combination of tones. Generalization of the classifier was evaluated using *k*-fold cross-validation with a 9:1 training to test ratio. The reported accuracy was the average across all 10 cross-validation folds. A sliding window was used to train and test each classifier, resulting in a time-varying measure of decoding accuracy. The length of the sliding window was 5 samples (corresponding to 50 ms of neural activity). As such, tones were decoded not only from their spatial activation patterns at each moment, but also the local temporal structure of responses within a given window. Classifier performance at each time point was evaluated in terms of balanced accuracy (Grootswagers et al., 2017), whereby accuracy was first determined individually for each class and then averaged across both.

Representational similarity analysis (RSA). The application of the above MVPA procedure to every pairwise combination of probe tones resulted in a 12 × 12 diagonally symmetric representational dissimilarity matrix (RDM) for every subject and time sample. These neural RDMs were compared with various model RDMs that evaluated the predictions of several perceptual and sensory models of pitch. Model RDMs were as follows: (1) An RDM based on the STH was constructed in which each cell coded the difference in perceived stability between two tones using the ratings first reported by Krumhansl and Kessler (1982). We considered the predictions of other perceptual models (Lerdahl, 1988, 2004; Chew, 2000); however, these resulted in RDMs that shared an identical rank-order structure to that of the STH RDM (i.e., the predicted dissimilarities between tones were the same as the STH) and were therefore omitted from the current analysis. (2) To test the hypothesis that MEG dissimilarities reflected the log-difference in each tone's fundamental frequency (f_0), we constructed a pitch height (PH) RDM, in which each cell corresponded to the semitone interval separating the two tones in question. (3) To assess whether neural dissimilarities between tones reflected their fine-grained spectral differences, a spectral distance RDM was constructed in which each cell coded the Euclidean distance between the 128-channel stimulus spectrograms of tones. Spectrograms were extracted by passing the raw audio through a biologically inspired model of the auditory periphery (Chi et al., 2005). The model consisted of three main stages: a cochlear filter bank comprised of 128 log-spaced asymmetric filters, a hair cell stage consisting of a low-pass filter and nonlinear compression function, and a lateral inhibitory network modeled as a first-order derivative along the tonotopic axis followed by a half-wave rectifier. (4) Although the tonal context and probe tones were separated by 650 ms (see Experimental design), models of auditory short-term memory involve time constants of up to 4 s (Huron and Parncutt, 1993; Leman, 2000). Thus, it was possible that neural dissimilarities between tones were driven by sensory memory of the context. To test this possibility, we constructed a spectral overlap (SO) RDM. First, a spectrogram was generated for the tonal context using the same model as that used for probe tones. Next, to account for the temporal decay of distal events, we exponentially weighed each frame of the context spectrogram such that recent frames had the strongest bearing on the spectral decomposition, using a time constant equal to 1 beat (tau = 0.65 s). We then calculated the Euclidean distance between context and all probe tone spectrograms. The differences in these spectral distances for each pairwise combination of probe tones were used to code each cell of the SO RDM. (5) Finally, to assess the potential impact of the in-key versus out-of-key identification task on neural dissimilarities, we coded a "task category" RDM (see Fig. 4*A*), whereby tone pairs belonging to the same decision category (either both in-key or both out-of-key) were coded with 0, whereas those differing in categorical membership were coded with 1. Using the framework of RSA (Kriegeskorte et al., 2008), we studied the brain's emerging representation by comparing each model RDM with the empirical time-varying neural RDM (see Statistical analysis).

Control experiment

Participants. Eleven participants were recruited to participate in a subsequent control experiment. The eligibility criteria and recruitment methods were identical to those used previously. One participant did not complete the testing session, rendering the dataset unusable. The 10 remaining participants had a mean of 10.0 years of musical training (SD = 3.7 years).

Experimental design. Trials consisted, as before, of a four-chord context followed by 1 of 12 possible probe tones, and the experiment was again structured into an hour-long session comprised of 8 blocks with 84 trials in each. Two manipulations were introduced into the control condition. First, to assess the impact of the behavioral task on neural decoding, participants were now uninformed of the experiment's purpose as it relates to pitch and instructed to perform a timbre-identification task instead, thus requiring no explicit analysis of pitch. In 12 trials per block (~15%), probe tones occurred with the timbre of a flute instead of piano. Upon hearing flute probe tones, subjects registered a button-press response. Such trials were randomly dispersed but constrained to occur exactly once on every pitch class per block, thereby preserving equal numbers of remaining piano tone exemplars within each pitch class. Second, to evaluate the influence of the distribution of pitch in the context on the neural processing of probe tones, two different versions of the context were presented in blocked design. In odd-numbered blocks, trials featured the original context chords voiced in 4-part harmony (cntx4). In even-numbered blocks, however, chords were voiced in 3-part harmony (cntx3) by removing the pitches [C4, C4, B3, C4] from the four chords, respectively (see Fig. 5*D*). In pilot testing, we found that this alteration sufficiently modified the statistical distribution of pitch without altering the percept of tonality. Waveforms corresponding to the two different contexts were then normalized for perceived loudness using the procedure previously used on probe tones (see Stimuli). All other aspects of the stimulus and design were identical to those used previously.

Analysis. Preprocessing followed an identical procedure to that used earlier with two exceptions. First, neural data corresponding to trials in which flute tones occurred were discarded from the analysis, yielding 48 remaining trials per pitch class. Second, we anticipated a poorer classification signal-to-noise ratio relative to the earlier analysis due to the lower number of subjects and trials. To combat this issue, continuous MEG timeseries were downsampled to 50 Hz before performing PCA (whereas 100 Hz was used previously). This lower temporal resolution was still sufficient as an examination of temporal dynamics was not paramount to the aims of the control experiment. MVPA parameters were also identical to the previous analysis, with the exception that a "leave-one-out" cross-validation scheme was adopted to maximize use of the data, and trial-averaging before classification was omitted when decoding separately across the two different contexts. To determine whether the statistical distribution of pitch in the preceding context explained the structure of dissimilarities between probe tones, we first coded candidate RDMs in which cells indexed pairwise differences in the statistical likelihood of tones within the context. Separate candidate RDMs were coded for cntx4 and cntx3, and we created two versions: one in which cells coded the

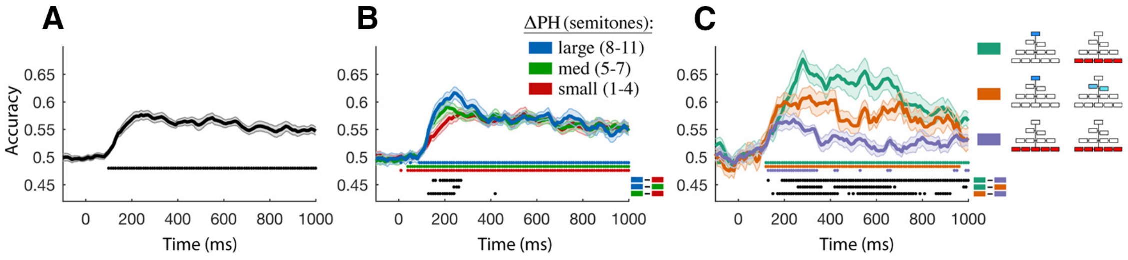

Figure 2. Temporal decoding of tones from MEG responses. The time domain in all plots are aligned such that 0 indicates onset of tones. ***A***, Average classification accuracy for decoding all pairwise combinations of the 12 tones. ***B***, Average classification accuracy when decoding tone pairs binned into three groups based on their PH separation: large (8–11 semitones; blue), medium (5–7 semitones; green) and small (1–4 semitones; red). ***C***, Classification of tone pairs grouped based on their difference in the hierarchy of perceived stability: large (green), medium (orange), and little-to-no difference (purple). Colored boxes in the schematic legend represent the hierarchical position of tones being decoded for each curve, with blue and red boxes representing stable and unstable tones, respectively. ***B***, ***C***, Results are averaged across all appropriate pairwise combinations of tones. Colored markers underneath curves represent time points when decoding performance differs significantly from chance levels ($p < 0.05$; Wilcoxon sign-rank tests, FDR-corrected). Black markers represent time points during which two decoding curves, specified by the bottom right colored boxes, are significantly different from one another. Shaded regions represent SEs across all participants ($N = 18$).

difference in pitch class probability for a given pair of probe tones, whereas another coded their f_0 probability difference instead. We then computed difference RDMs (ΔRDMs) by subtracting RDMs corresponding to cntx3 from that of cntx4. Finally, we examined correlations between the resulting neural and candidate ΔRDMs (see Fig. 5*H*).

Adaptation to regularities across experiment. To assess whether the structure of neural RDMs reflected gradual adaptation of MEG responses to stimulus regularities over the course of the entire eight-block experiment, we performed MVPA repeatedly on four-block windows of adjacent data, proceeding in single-block steps from the first to last block. We correlated each of the resulting neural RDMs (see Fig. 6*A*) to three different candidate RDMs (see Fig. 6*B*): SO, f_0 probability, and pitch class probability. Each candidate RDM was coded as previously, using a uniform temporal weighting of the context distribution in each case.

Temporal generalization of classifiers. To evaluate the dynamics of activation patterns distinguishing different probe tones, we examined the temporal generalization of classification. Classifiers trained at individual time points were tested at every other time point, resulting in a square temporal generalization matrix (see Fig. 7*A*). A classifier trained at a given time t whose performance generalizes to another time t' implies that the neural code separating tones at t recurred at t'. In this fashion, insight can be gained into the underlying dynamics of neural information processing (King and Dehaene, 2014). Classification parameters were identical to those used previously, with the exception that the window size was reduced to a single time sample. Temporal generalization matrices were averaged across all pairwise combinations of tones. To quantify the extent of generalization, we calculated the spread of decoding for each train time (i.e., each horizontal slice of the temporal generalization matrix) at which any significant decoding occurred. Spread was operationalized as the full-width of the slices at 75% of peak decoding. Varying the percentage at which this width was calculated did not alter trends.

Statistical analysis. Significance of classification was performed at the group level ($N = 18$) at each time sample and was evaluated using two-sided Wilcoxon sign-rank tests ($p < 0.05$). Multiple comparisons were corrected by controlling the false discovery rate (FDR) (Benjamini and Yekutieli, 2001; Nichols, 2012) with $\alpha = 0.05$. All correlations between RDMs were assessed by computing their rank-order Kendall's T_A, a measure that allows for the comparison of continuous and categorical matrices (Nili et al., 2014). Time-varying RDM correlations were FDR-corrected, whereas time-averaged RDM correlations were Bonferroni-Holm–corrected (Aickin and Gensler, 1996). We used the noise ceiling as a benchmark for testing candidate RDM performance. The noise ceiling uses the intersubject variance in RDMs to estimate the magnitude of the expected correlation between a "true" model RDM and the empirical RDM, given inherent measurement noise (Nili et al., 2014). To compare the predictive capacity of two different models, paired Wilcoxon sign-rank tests were performed on time-averaged correlations. Such comparisons were only made between candidate RDMs that offered significant predictive power at any time point in the neural epoch. To visualize the structure of cortical RDMs, multidimensional scaling (MDS) was applied using Kruskal's normalized stress 1 criterion (Kruskal and Wish, 1978). The correlation between individual subject RDMs (see Fig. 3*B*) was calculated by averaging the Kendall's T_A correlation across all pairwise combinations of subjects' RDMs, and significance was assessed using FDR-corrected permutation tests (Nichols, 2012). Significance of adaptation over the course of the experiment was assessed by applying a one-way ANOVA to the neural-candidate RDM correlations across early to late windows of neural data (see Fig. 6*C*) and evaluating the main effect of time.

Results

Results are derived from cortical MEG recordings during the presentation of 12 different "probe tones" that spanned the set of all pitch classes within an octave following a C major tonal context (see Materials and Methods). Classifiers attempted to decode the identity of two given probe tones from their neural activity at each time sample, and the resulting decoding accuracy provided a time-varying estimate of the dissimilarity in their cortical population codes. Applying this procedure to every pairwise combination of the 12 different probe tones, we characterized the dynamic representational structure of musical pitch in cortex.

Neural response patterns initially code the f0 of tones but later code their perceived stability

We first examined the general dynamics of stimulus-specific information in cortex, assessing the average decoding performance across all pairwise combinations of tones (Fig. 2*A*). As expected, accuracy was at chance (50%) before the onset of tones ($t = 0$) as stimulus-related information was yet to activate cortex. Neural distinctions between tones first emerged 100 ms after onset, reached a maximal value at 250 ms, and remained above chance for the remainder of the epoch.

We next examined the dissimilarity between tones whose acoustic or perceptual properties generate explicit predictions regarding their representational distance. First, as tones differed from one another in the periodicity of their waveforms, we reasoned that their distinctions in cortex may be commensurate with their log-f_0 separation, which we term PH. Decoding performance was therefore examined for pairwise combinations of tones grouped based on whether their PH difference was small (1–4 semitones), medium (5–7 semitones), or large (8–11 semitones). Indeed, we found that the magnitude of PH separation between tones produced differences in decoding performance from ~100 to 250 ms after onset (Fig. 2*B*). Cortical distinctions between tones with large and medium PH differences (blue and green curves, respectively) significantly exceeded those between tones that had small PH separation (red curve).

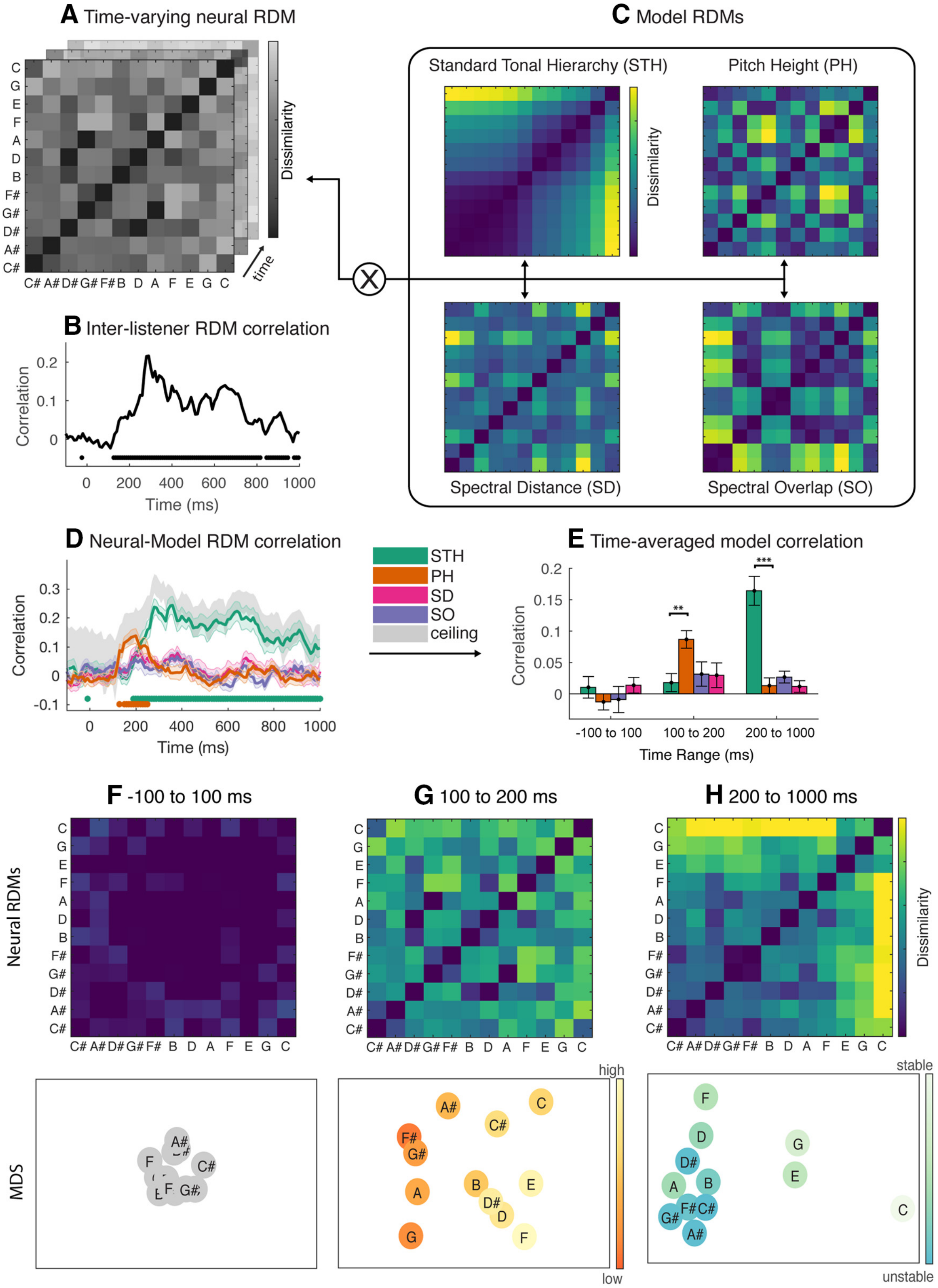

Figure 3. RSA of musical pitch. ***A***, Neural representational similarity matrix (RDM) indicating the MEG dissimilarities between pairs of pitch classes at each time point in the neural epoch. ***B***, Mean rank-order correlation between the neural RDMs of individual listeners ($N = 18$). Significant time points are indicated underneath the curve ($p < 0.05$; randomization test, FDR-corrected). ***C***, Four different candidate RDMs based on models that attempt to explain neural dissimilarities. ***D***, Rank-order correlations between each model RDM and neural RDMs at every time point. Shaded regions represent SEs across listeners. Colored markers beneath curves represent significant time points ($p < 0.05$; Wilcoxon sign-rank tests, FDR-corrected). ***E***, For visualization purposes, neural-model RDM correlations were averaged across three different peristimulus time bins. ***F–H***, Average neural RDMs (top) and MDS solutions (bottom) for the three different time bins in ***E***. Colormaps represent PH (low to high) or perceived stability (unstable to stable) in ***G*** and ***H***, respectively. ***E***: ** indicates $p < 0.01$; *** indicates $p < 0.001$.

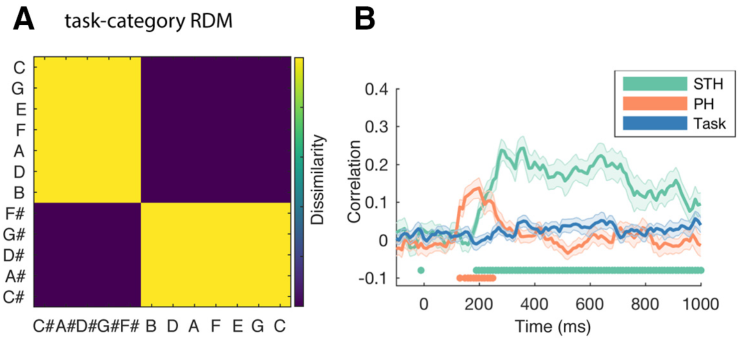

Figure 4. Model testing of task-related dissimilarity between tones. ***A***, Binary "task category" candidate RDM based on whether each pairwise combination of tones is the same (0) or different (1) in the in-key versus out-of-key distinction. ***B***, Correlation between task category and neural RDMs (blue). STH and PH correlations are also shown for reference. Shaded regions indicate SEs across listeners.

Second, tones differed in their perceived stability given the preceding musical context. We therefore hypothesized that distinctions in their cortical encoding may honor these perceptual differences, embodied by the STH. If this were true, MEG decoding performance would be greatest for tones located at opposite ends of the hierarchy and poorest for tones that are hierarchically equal. Support for this hypothesis was observed at an epoch that commenced ∼200 ms after onset and persisted throughout the duration of the epoch (Fig. 2*C*). Neural responses to the most stable tone [C] were highly distinct from those of the unstable tones [F#, G#, D#, A#, C#] (green curve) and less discriminable from those of the second and third most stable tones [G and E, respectively] (orange curve). Additionally, consistent with the STH, classifiers performed poorly when attempting to distinguish the neural activity of unstable tones from one another (purple curve). These results suggest that the extent to which the cortical activity elicited by two tones differ corresponds to the difference in their position within the STH.

Together, these results suggest that early cortical distinctions between tones reflect their absolute pitch (log-f_0) differences, whereas later distinctions reflect their perceived stability given the preceding musical context, as modeled by the STH. We next tested this hypothesis explicitly within the framework of RSA (Kriegeskorte et al., 2008). For each subject, we indexed the set of decoding accuracies obtained by classifying all pairwise combinations of tones in a time-varying RDM (Fig. 3*A*). Each cell of the diagonally symmetric RDM indicates the cortical dissimilarity between the tones indexed by the respective row and column. We found that RDMs were highly correlated across individual subjects from 100 ms onwards (Fig. 3*B*), verifying that the representational structure was consistent across listeners over the same temporal extent in which average stimulus distinctions were apparent (Fig. 2*A*).

Next, we evaluated the predictive capacity of several models that attempt to explain the observed structure of time-varying cortical RDMs. Each model was coded as a candidate RDM whose structure makes explicit predictions regarding the expected dissimilarities between tones (Fig. 3*C*). One candidate RDM was based on the STH, where distances between tones corresponded to their difference in perceived stability, as reported by Krumhansl and Kessler (1982). Another candidate RDM coded for differences in PH to evaluate whether distinctions between tones were driven by their log-f_0 separation. We additionally tested two purely sensory models: one based on the spectral distance between tone pairs, and another based on the differences in their SO with the preceding musical context (see Materials and Methods). Each subject's cortical RDM at every time point was compared with the four-different candidate RDMs using a rank-order correlation measure, resulting in four curves tracking neural-model correlation across time (Fig. 3*D*). Consistent with the implications of our earlier analysis, the PH and STH models significantly accounted for the variance in cortical RDMs in early (100–250 ms) and later (190 ms onwards) regions of neural processing, respectively. Additionally, directly comparing the predictive power of the two models, we found that PH performed significantly better than STH in the earlier period ($Z = 2.74$; $p = 0.003$), whereas STH significantly outperformed PH during the latter period ($Z = 3.35$; $p = 3.99 \times 10^{-4}$). Importantly, both PH and STH correlations closely tracked the noise ceiling (Nili et al., 2014), indicating that these models offered an optimal degree of predictive power given the level of noise inherent in the MEG data (see Materials and Methods). The temporal order of model correlations is consistent with dominant conceptions of hierarchical auditory processing, which posit the extraction of complex pitch before the integration and analysis of broader tonal-harmonic structure (Koelsch, 2011). Interestingly, from 190 to 250 ms, PH and STH models were both significantly correlated with cortical RDMs, suggesting an intermediary period during which the cortex holds a combined representation of both the tone's f_0 and tonal status within the STH.

To better visualize the results of RSA, neural-model RDM correlations were averaged into three time bins (Fig. 3*E*): the first corresponded to a period before stimulus-specific information was present in cortical activity (−100 to 100 ms); the second corresponded to the period during which cortical structure was most strongly correlated with PH differences (100 to 200 ms); and the third corresponded to the remainder of the neural epoch, during which cortical structure reflected the STH (200–1000 ms). Time-averaged neural RDMs corresponding to each of the three bins are displayed in Figure 3*F–H* (top). To more intuitively visualize their dissimilarity structure, we applied MDS to each RDM, obtaining a 2D solution in each case (Fig. 3*F–H*, bottom). The MDS solution in Figure 3*G* clearly demonstrates the organization of pitch from low to high as the space is traversed from top left to bottom right, respectively. Similarly, the spatial organization of the MDS solution in Figure 3*H* illustrates many key properties of the STH. Traversing the space from right to left reveals the structure of the hierarchy, with the most stable pitch class (C) situated on the right side, closest to the next most stable classes (G and E) but distant from the cluster of unstable classes (F#, G#, D#, A#, C#) in the bottom left corner. Prior behavioral research has underscored the perceptual primacy of this hierarchical arrangement. Our findings now provide evidence of its origins in the cortex and reveal the temporal dynamics with which it emerges from the acoustic signal via an intermediate representation of PH.

Emergence of tonal hierarchy in cortex cannot be explained by activity generated from task-related judgments

During the experiment, subjects were instructed to identify whether probe tones were in-key or out-of-key given the preceding context (see Materials and Methods). As this decision variable was closely related to the perceptual dimension being decoded (the STH),

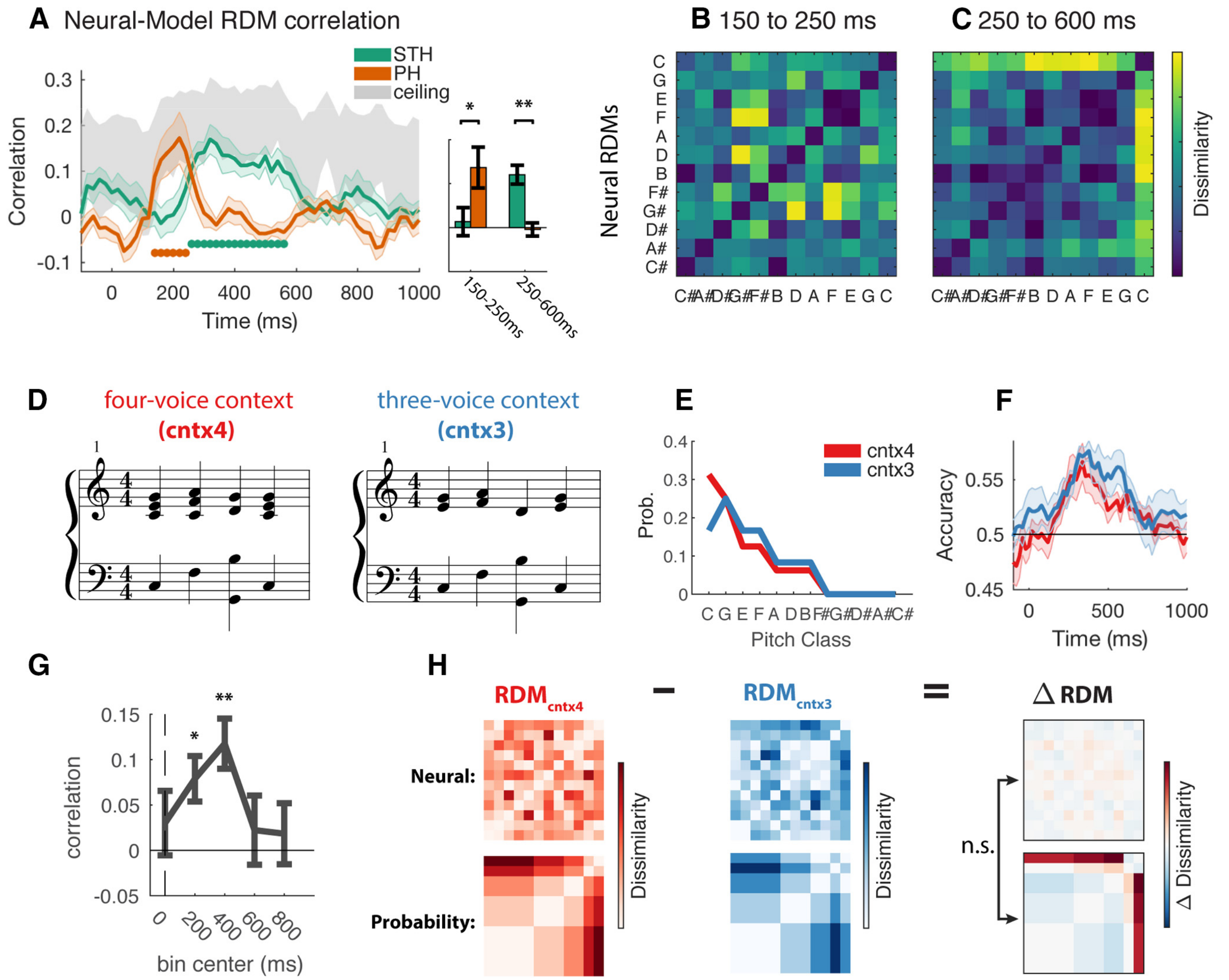

Figure 5. Results of control experiment. ***A***, Left, Correlations between PH and STH model RDMs and neural RDMs at every time point. Gray shaded region represents the correlation noise ceiling. Right, Time-averaged correlations across the two regions during which PH and STH model correlations, respectively, were significantly above baseline. ***B***, ***C***, Neural RDMs averaged across the same two temporal regions. ***D***, Musical score corresponding to the four-voice context (cntx4; see Figure 5-1, available at https://doi.org/10.1523/JNEUROSCI.1399-19.2020.f5-1; and Figure 5-2, available at https://doi.org/10.1523/JNEUROSCI.1399-19.2020.f5-2) and three-voice context (cntx3; see Figure 5-3, available at https://doi.org/10.1523/JNEUROSCI.1399-19.2020.f5-3; and Figure 5-4, available at https://doi.org/10.1523/JNEUROSCI.1399-19.2020.f5-4). ***E***, Probability distribution of pitch class in cntx4 and cntx3. ***F***, Group mean decoding accuracy for discriminating MEG responses of the most stable tone [C] from the set of tones [E, F, A, D, B, F#, G#, D#, A#, C#] when presented within cntx4 and cntx3. ***G***, Correlation between neural RDMs in cntx4 and cntx3 averaged into 200 ms bins. ***H***, Neural ΔRDM (top row) and candidate ΔRDM based on either pitch class probability (bottom row) or f_0 probability (data not shown). ΔRDMs were computed by subtracting RDMs corresponding to cntx3 from cntx4. *$p < 0.05$. **$p < 0.01$. Shaded regions indicate SEs across subjects ($N = 10$).

neural decoding may have been driven by activity corresponding to the maintenance of a decision in working memory during the span of the neural epoch. To address this issue, we tested an additional candidate RDM that coded the binary dissimilarity structure of the behavioral task (Fig. 4*A*). Specifically, tone pairs belonging to the same decision category (either both in-key or both out-of-key) were coded with 0, whereas those differing in this categorical membership were coded with 1. This candidate RDM did not correlate significantly with neural RDMs at any time during the epoch (Fig. 4*B*), indicating that decision-related variance could not account for the neural representational structure between tones.

Even though the categorical membership of tones as in-key or out-of-key was not directly evidenced in the structure of neural RDMs, it remained possible that the steering of listeners' attention to a perceptual dimension requiring pitch analysis influenced the geometry of neural RDMs in complex ways not captured by simple linear modeling. Therefore, in a control condition conducted on a separate cohort of musically trained subjects, we determined whether the STH can still predict neural dissimilarities between tones even under conditions where attention is directed to a dimension outside of pitch. As before, subjects were presented with probe tones following a context during MEG recordings; however, they were now uninformed of the experiment's purpose as it relates to pitch, and performed a timbre-identification task instead (for details, see Materials and Methods). Although evidence suggests that the cortical processing of tonal schema is modulated by attention (Loui et al., 2005), we hypothesized the involvement of an automatic component, whereby listeners implicitly encode the structure of the STH even in the absence of a task requiring them to do so.

Consistent with the above hypothesis, we found that representational dynamics in the control experiment closely mimicked those found originally (Fig. 5*A*). Specifically, in an early temporal

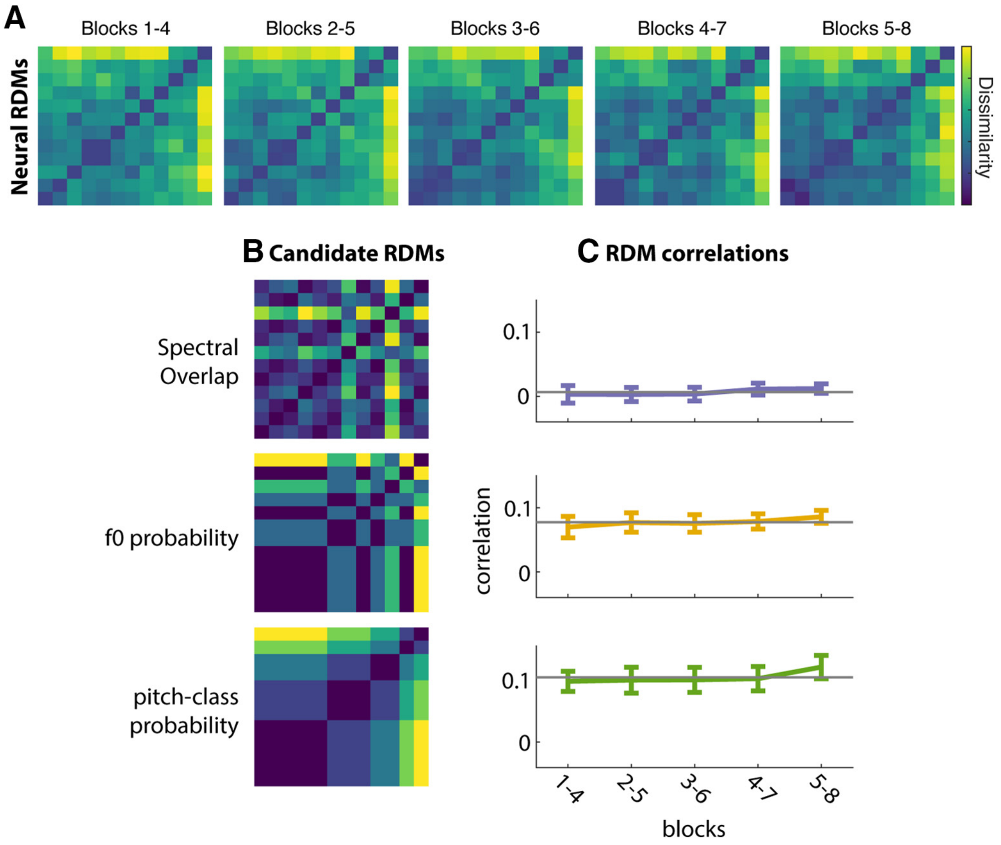

Figure 6. Representational structure over the course of experiment. ***A***, Neural RDMs resulting from classification of subsets of MEG data spanning early to late periods of the experiment. ***B***, Candidate RDMs based on pairwise differences in statistical regularity at three levels of stimulus representation: spectral (top), f_0 (middle), and pitch class (bottom). ***C***, Neural-candidate RDM correlations across early-to-late windows for the corresponding models in ***B***. Gray lines indicate the mean correlation value across all five windows. Error bars indicate SEs across participants ($N = 18$).

region from 150 to 250 ms, PH was significantly correlated with neural RDMs and outperformed the STH model ($Z = 1.73$; $p = 0.04$), whereas a subsequent broader period (250–600 ms) emerged during which the STH predicted neural dissimilarities and significantly outperformed the PH model ($Z = 2.55$; $p = 0.005$). Accordingly, the time-averaged neural RDMs from these two respective periods (Fig. 5*B*,*C*) closely resemble those found earlier (Fig. 3*G*,*H*). These results verify that the cortical representation of musical pitch emerges independent of activity generated from task-related judgments.

Cortical emergence of tonal hierarchy cannot be explained by stimulus regularities within the experiment

Our central hypothesis, that neural dissimilarities reflect the tonal hierarchy, is presumed to arise from an internal schema acquired through exposure to the structure of tonal music across the lifetime. However, as probe tones were repeatedly presented within a fixed context, neural responses may have been driven by the structural regularities existing within the experiment itself, reflecting adaptation or statistical tracking of the stimulus. Therefore, we next examined whether neural dissimilarities between tones could be explained without invoking any prior knowledge, by the cortical formation of a representation of recent stimulus information contained strictly within the bounds of the experiment. If findings reflect a genuine perceptual schema, the geometry of neural RDMs should be unaffected by bottom-up relationships between probe tones and the acoustic or statistical structure of the experiment.

First, we considered whether the structure of the preceding context on each trial could explain neural dissimilarities between the subsequent probe tones. A candidate RDM based on differences in SO between probe tones and context failed to predict cortical dissimilarities (Fig. 3*D*), suggesting that MEG decoding was not driven by adaptation to the prior context at a sensory level. However, listeners are also known to track the statistics of higher-level properties, including f_0 and pitch class, during perception of tonal sequences (Saarinen et al., 1992; Saffran et al., 1999; Creel et al., 2004). In the current design (as in most tonal music), the probability distribution of pitch classes in the constituent chords of the context (Fig. 5*E*, red curve) closely tracked the profile of the STH itself (Fig. 1). Consequently, those pitch classes with highly distinct neural activation patterns (e.g., C) were also the most frequently occurring. This raises the possibility that, rather than a genuine perceptual schema acquired through long-term exposure, neural RDMs may index differences in surprisal between probe tones given the statistical context in which they occur.

To test whether neural dissimilarities between tones were influenced by the frequency of their occurrence in the prior context, subjects in a control condition were presented with two variants of the context (Fig. 5*D*). Half of the trials contained the original chords voiced in 4-part harmony (cntx4), whereas the other half contained chords with the most probable pitches removed (see Materials and Methods). This resulted in an alternate 3-voice chord progression (cntx3) that still unambiguously established the key of C major but crucially disrupted the association between the statistical likelihood of each pitch class and its position within the STH (Fig. 5*E*). We hypothesized that a stable internal knowledge of tonal structure underlies the observed structure in the brain; and as such, neural RDMs would remain invariant to statistical manipulations of the context. In particular, the likelihood of the most perceptually stable pitch class [C] differed by an approximate factor of 2 across the two contexts. Furthermore, in terms of absolute pitch (f_0), the tone C4 was highly probable within cntx4 but completely absent from cntx3. If neural responses reflect adaptation or tracking of short-term statistical regularities, evoked activity to C should be less distinct from the other probe tones when preceded by cntx3. On the other hand, if neural responses reflect the perceived musical stability of tones, activity evoked by C should remain equally distinct across both contexts. Separating probe tone responses that were preceded by cntx4 and cntx3, and independently assessing MEG decoding performance in each instance, we found that classifiers were able to discriminate responses to C from the other probe tones equally well in both contexts (Fig. 5*F*; across all time points: $Z < 0.97$; $p > 0.62$), thereby supporting our hypothesis. To examine neural dissimilarities across the two contexts in more detail, we next generated two independent sets of time-varying neural RDMs by separately decoding all pairwise combinations of probe tones preceded by cntx4 and cntx3. We assessed their relationship to one another by examining RDM correlations averaged into 200 ms bins (Fig. 5*G*). RDMs across cntx4 and cntx3 were significantly correlated in two bins centered at 200 ms ($Z = 2.45$; $p = 0.0195$) and 400 ms ($Z = 2.75$; $p = 0.005$) after onset. Last, we sought to systematically determine whether differences, if any, in the structure of neural RDMs across cntx4 and cntx3 could be explained by differences in the relative pitch class probability of probe tones. We first computed a neural difference RDM (ΔRDM) by subtracting the time-averaged cntx3 RDM from that of cntx4 (Fig. 5*H*, top row). Similarly, we constructed a candidate ΔRDM that coded differences in the relative probability of each pitch class across cntx3 and cntx4 (Fig. 5*H*, bottom row). We found that neural and candidate ΔRDMs were not significantly correlated (mean $T_A = -0.05$; $Z = -1.42$; $p = 0.92$), suggesting that any variation in the structure of neural RDMs between cntx4 and cntx3 could not be accounted for by statistical differences between the two contexts at the level of pitch class. Modeling the differences in statistical regularity in terms of absolute pitch (f_0) instead of pitch class also failed to explain neural RDM differences between cntx4 and cntx3 (mean $T_A = -0.09$; $Z = -1.94$; $p = 0.97$). Finally, to account for temporal recency effects, we exponentially weighed the pitch class and f_0 distributions of both contexts such that constituent tones that were temporally proximate to probe tones received greater weights. Three different weighted pitch distributions using exponential time constants of 1, 2, and 3 beats, respectively (1 beat = 650 ms) all failed to produce significant correlations between candidate and neural ΔRDMs, regardless of whether distributions were modeled on pitch class or f_0 (all $Z < -1.43$; $p > 0.92$). Together, these results indicate that neither the low-level acoustic nor the high-level statistical structure of the preceding context can account for neural dissimilarities between tones.

The above findings suggest that the STH emerges in cortex independently from the specific structure of the context that immediately preceded tones. However, because the hour-long experiment used a fixed subset of pitches, adaptation may have operated at longer timescales than the duration of a single instance of the context. To this end, we next examined whether the geometry of neural RDMs changed over the course of the entire experiment in a manner that reflects gradual adaptation to regularities in the structure of the repeatedly occurring context. For each subject's dataset, we again constructed neural RDMs by applying MVPA: classifying the neural activity corresponding to every pairwise combination of probe tones. However, we now only used MEG data from four adjacent experimental testing blocks (of eight total), and repeatedly applied MVPA using a sliding window proceeding in single block steps. This resulted in five distinct windowed neural RDMs reflecting the cortical structure between tones at progressively later windows during the experiment (Fig. 6*A*). Comparing correlations between each windowed neural RDM (time-averaged from 200 to 1000 ms) and the STH, we found no difference in the extent to which the tonal hierarchy was reflected in neural data collected earlier versus later in the experiment (one-way ANOVA; $F_{(4,80)} = 0.17$; $p = 0.95$). To directly test whether MEG activity at later periods increasingly reflected the acoustic or statistical regularities inherent in the experiment, we correlated each windowed neural RDM with three candidate RDMs that each predict dissimilarities between

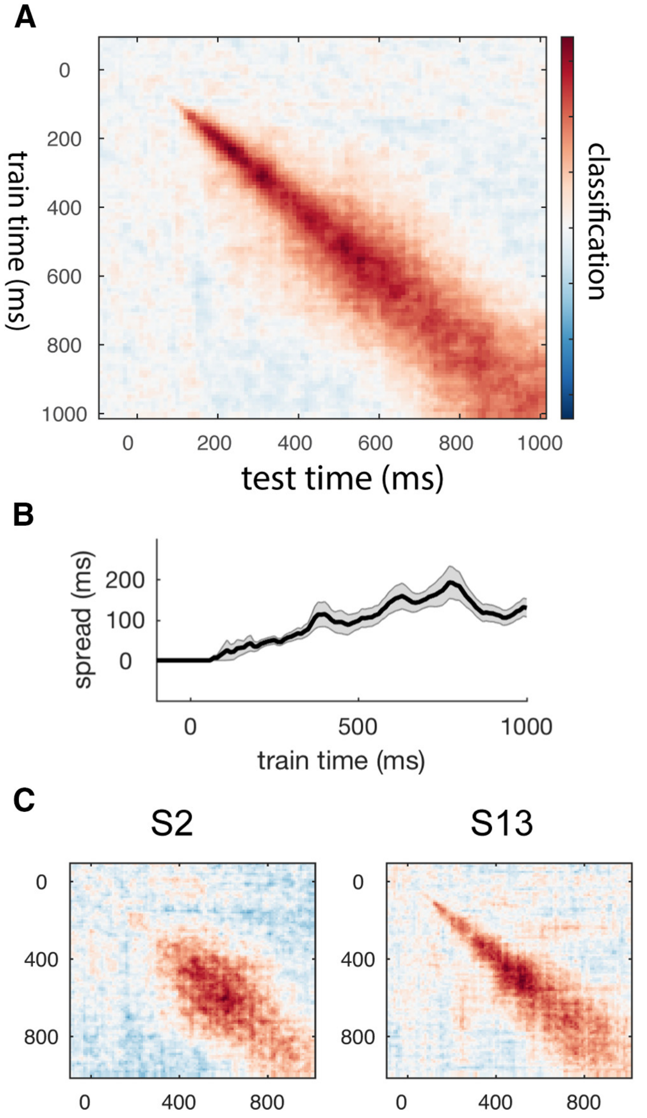

Figure 7. Temporal generalization of classifiers. ***A***, Temporal generalization matrix averaged across all subjects and pairwise combinations of probe tones. ***B***, Spread of generalization across train time points. Shaded regions indicate SEs across subjects ($N = 18$). ***C***, Temporal generalization matrices for 2 individual subjects.

tone pairs based on the regularity of their occurrence in the context at different levels of representation (Fig. 6*B*). If activity increasingly reflected structural regularities in the unfolding experiment, correlations with one or more candidate RDMs should monotonically increase as a function of experiment time. Importantly, however, correlations remained unchanged as neural data were taken from progressively later trials (Fig. 6*C*; one-way ANOVA, $F_{(4,80)} < 0.24$; $p > 0.91$ for all models), indicating that stimulus adaptation across timescales spanning the entire experiment did not have a significant bearing on the cortical representation of tonal structure.

In sum, we show that the representation of musical pitch in cortex cannot be accounted for by purely bottom-up processes that posit the integration of (or adaptation to) experimental stimulus structure at varying representational levels or timescales. Instead, the structure of neural RDMs >200 ms reflects a stable mental representation of hierarchical pitch structure, acquired through a lifetime of exposure to the structure of tonal music.

Cortical population codes become increasingly stable with time

Having characterized the representational dynamics of pitch throughout the time course of neural processing, we last examined the dynamics of the underlying neural patterns that code this information. Electrophysiological recordings during melodic perception indicate that evoked components associated with initial acoustic feature analysis are relatively transient, whereas later components associated with tonal-harmonic analysis and contextual integration are activated across broader temporal extents (Besson and Macar, 1987; Besson and Faita, 1995; Koelsch, 2009, 2011). We therefore hypothesized that patterns of cortical activity would become increasingly stable across time as the information being coded transitioned from a continuous physical stimulus feature to a more discrete perceptual object. To test this hypothesis, we evaluated the capacity of classifiers trained at a given time point to generalize their performance across time (King and Dehaene, 2014), resulting in a train time × test time generalization matrix (Fig. 7*A*). In the first 200 ms after onset of tones, classifiers could only decode information when trained and tested on activity from the same time points, evidenced by the concentration of decoding performance along diagonal cells of the matrix. The lack of generalization during this initial period suggests that the neural code separating tones was dynamically evolving rather than stable. Interestingly, however, the generalization of classifiers improved across time, evidenced by an increase in off-diagonal decoding at later train times. We further verified this effect by calculating the spread of decoding at each train time, finding a monotonic increase in generalization across time (Fig. 7*B*). Although the specific profile of generalization varied across individual subjects (Fig. 7*C*), the overall trends confirm our hypothesis that populations encoding musical pitch have increasingly stable dynamics with time. Interestingly, together with earlier findings detailing the representational dynamics of pitch (Fig. 3*D*), this result suggests that the sensory-to-perceptual transition in information coding is accompanied by a transition in cortical dynamics, from transient to stable.

Discussion

Using temporal decoding techniques, we have characterized the representation of the 12 chromatic pitch classes of Western tonal music in human cortex. Our key finding, that dissimilarities in the cortical population codes between tones correspond to their differences in perceived stability, establishes the neurobiological reality of the tonal hierarchy: the first known study to comprehensively do so. However, consistent with a hierarchical view of auditory processing (Koelsch, 2011), we also found that the transformation of tones from acoustic waveforms to discrete perceptual entities is mediated by at least one intermediate representational level, during which cortical coding reflects the dimension along which tones are ordered from low to high by their f_0 value.

Although we have examined the representation of tones within one musical key, it should be noted that the second-order representational structure that exists between different musical keys naturally emerges as a consequence. Specifically, transposing the STH into different keys and correlating each profile with one another, the theoretical key relations comprising the "circle-of-fifths" have been empirically recovered (Krumhansl and Kessler, 1982). Thus, because we have detailed the encoding of the specific structure of the STH in cortex, we also establish the neurobiological basis of the structure among different keys. This represents a significant advance in understanding the neural correlates of tonal structure. Prior research examining key relationships have used fMRI (Janata et al., 2002); and limited by its relatively poor temporal resolution, these measurements reflect the general accumulation of information across an extended musical passage rather than mechanisms operating at the timescale of an individual tone.

Having established the basis of musical pitch representations and the dynamics with which they emerge in cortex, several outstanding questions remain. First, what are the underlying computations performed by populations in the auditory pathway that transform low-level stimulus features, such as basic frequency information, into a percept highly abstracted from acoustics? Elucidating the architecture and mechanisms of such a network is relevant to the more domain-general question of how incoming sensory input gets integrated with schematic "top-down" knowledge, a central problem in the study of human perception. Second, while current findings are based solely on data collected from musically trained subjects, future work should examine whether results generalize to nonmusicians. More broadly, to address outstanding questions relating to the effects of domain expertise on perception, research should explore how the neural substrates of tonal structure emerge in the auditory system throughout the course of musical development. Computational modeling has shown that simple networks exposed to the natural statistics of tonal music are capable of developing hierarchical pitch structure (Tillmann et al., 2000), suggesting that the brain may heuristically learn such representations over the lifetime. The current paradigm should therefore be extended to include subjects at various levels of musicianship to map the time course and examine specific mechanisms associated with learning of pitch structure.

References

Aickin M, Gensler H (1996) Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. Am J Public Health 86:726–728.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 1165–1188.

Besson M, Faita F (1995) An event-related potential (ERP) study of musical expectancy: comparison of musicians with non-musicians. J Exp Psychol Hum Percept Perform 21:1278.

Besson M, Macar F (1987) An event related potential analysis of incongruity in music and other non-linguistic contexts. Psychophysiology 24:14–25.

Brattico E, Tervaniemi M, Näätänen R, Peretz I (2006) Musical scale properties are automatically processed in the human auditory cortex. Brain Res 1117:162–174.

Chew E (2000) Towards a mathematical model of tonality. Doctoral dissertation. Massachusetts Institute of Technology.

Chi T, Ru P, Shamma SA (2005) Multiresolution spectrotemporal analysis of complex sounds. J Acoust Soc Am 118:887–906.

Creel SC, Newport EL, Aslin RN (2004) Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. J Exp Psychol Learn Mem Cogn 30:1119 –1130.

Duda RO, Hart PE, Stork DG (2012) Pattern classification. New York: Wiley.

Fedorenko E, McDermott JH, Norman-Haignere S, Kanwisher N (2012) Sensitivity to musical structure in the human brain. J Neurophysiol 108:3289 –3300.

Foo F, King-Stephens D, Weber P, Laxer K, Parvizi J, Knight RT (2016) Differential processing of consonance and dissonance within the human superior temporal gyrus. Front Hum Neurosci 10:154.

Glasberg BR, Moore BC (2002) A model of loudness applicable to time-varying sounds. J Audio Eng Soc 50:331–342.

Griffiths TD, Büchel C, Frackowiak RS, Patterson RD (1998) Analysis of temporal structure in sound by the human brain. Nat Neurosci 1:422– 427.

Grootswagers T, Wardle SG, Carlson TA (2017) Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. J Cogn Neurosci 29:677– 697.

Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield AQ (2002) Spectral and temporal processing in human auditory cortex. Cereb Cortex 12:140 –149.

Haxby JV, Connolly AC, Guntupalli JS (2014) Decoding neural representational spaces using multivariate pattern analysis. Annu Rev Neurosci 37:435– 456.

Huron D, Parncutt R (1993) An improved model of tonality perception incorporating pitch salience and echoic memory. Psychomusicology 12:154.

Janata P, Birk JL, Van Horn JD, Leman M, Tillmann B, Bharucha JJ (2002) The cortical topography of tonal structures underlying Western music. Science 298:2167–2170.

Kado H, Higuchi M, Shimogawara M, Haruta Y, Adachi Y, Kawai J, Ogata H, Uehara G (1999) Magnetoencephalogram systems developed at KIT. IEEE Trans Appl Superconductivity 9:4057– 4062.

King JR, Dehaene S (2014) Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn Sci 18:203–210.

Klein ME, Zatorre RJ (2011) A role for the right superior temporal sulcus in categorical perception of musical chords. Neuropsychologia 49:878 – 887.

Koelsch S (2009) Music-syntactic processing and auditory memory: similarities and differences between ERAN and MMN. Psychophysiology 46:179 –190.

Koelsch S (2011) Toward a neural basis of music perception: a review and updated model. Front Psychol 2:110.

Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis: connecting the branches of systems neuroscience. Front Syst Neurosci 2:4.

Krohn KI, Brattico E, Välimäki V, Tervaniemi M (2007) Neural representations of the hierarchical scale pitch structure. Music Perception 24:281–296.

Krumhansl CL, Kessler EJ (1982) Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. Psychol Rev 89:334 –368.

Krumhansl CL, Shepard RN (1979) Quantification of the hierarchy of tonal functions within a diatonic context. J Exp Psychol Hum Percept Perform 5:579 –594.

Kruskal JB, Wish M (1978) Multidimensional scaling: Sage University Paper series on quantitative applications in the social sciences. Beverly Hills, CA: Sage.

Lee YS, Janata P, Frost C, Hanke M, Granger R (2011) Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. Neuroimage 57:293–300.

Leman M (2000) An auditory model of the role of short-term memory in probe tone ratings. Music Percept 17:481–509.

Lerdahl F (1988) Tonal pitch space. Music Percept 5:315–349.

Lerdahl F (1992) Cognitive constraints on compositional systems. Contemp Music Rev 6:97–121.

Lerdahl F (2004) Tonal pitch space. Oxford: Oxford UP.

Loui P, Grent T, Torpey D, Woldorff M (2005) Effects of attention on the neural processing of harmonic syntax in Western music. Cogn Brain Res 25:678 – 687.

Nichols TE (2012) Multiple testing corrections, nonparametric methods, and random field theory. Neuroimage 62:811– 815.

Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N (2014) A toolbox for representational similarity analysis. PLoS Comput Biol 10:e1003553.

Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. Neuron 36:767–776.

Saarinen J, Paavilainen P, Schöger E, Tervaniemi M, Näätänen R (1992) Representation of abstract attributes of auditory stimuli in the human brain. Neuroreport 3:1149 –1151.

Saffran JR, Johnson EK, Aslin RN, Newport EL (1999) Statistical learning of tone sequences by human infants and adults. Cognition 70:27–52.

Sankaran N, Thompson WF, Carlile S, Carlson TA (2018) Decoding the dynamic representation of musical pitch from human brain activity. Sci Rep 8:839.

Thwaites A, Glasberg BR, Nimmo-Smith I, Marslen-Wilson WD, Moore BC (2016) Representation of instantaneous and short-term loudness in the human cortex. Front Neurosci 10:183.

Tillmann B, Bharucha JJ, Bigand E (2000) Implicit learning of tonality: a self-organizing approach. Psychol Rev 107:885–913.

Uehara G, Adachi Y, Kawai J, Shimogawara M, Higuchi M, Haruta Y, Ogata H, Kado H (2003) Multi-channel SQUID systems for biomagnetic measurement. IEICE Trans Electronics 86:43–54.

Vanrullen R (2011) Four common conceptual fallacies in mapping the time course of recognition. Front Psychol 2:365.

Vos PG, Troost JM (1989) Ascending and descending melodic intervals: statistical findings and their perceptual relevance. Music Percept 6:383–396.

Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP (2001) Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. J Cogn Neurosci 13:1–7.

Zatorre RJ, Evans AC, Meyer E (1994) Neural mechanisms underlying melodic perception and memory for pitch. J Neurosci 14:1908 –1919.